Validación de un modelo basado en Técnicas de Evaluación Multicriterio y SIG a partir de análisis de incertidumbre. Una propuesta para mejorar la introducción de distorsión en las variables de partida

A. Calabia, M. Gómez y G. Benavidez

Departamento de Geografía. Universidad de Alcalá.

andres@calabia.com, montserrat.gomez@uah.es, gracebeg@hotmail.com

RESUMEN

Un modelo es siempre una versión simplificada de la realidad que nos permite describir y comprender mejor un determinado problema, recogiendo los elementos y mecanismos esenciales de los sistemas del mundo real. Sin embargo, un aspecto necesario es demostrar que esos instrumentos generan representaciones fiables de los sistemas que simulan a través de la aplicación de un proceso de validación.

Nuestro modelo a validar reproduce un escenario futuro de crecimiento urbano a partir de técnicas de Evaluación MultiCriterio (EMC) y Sistemas de Información Geográfica (SIG). El problema crucial en este ámbito es averiguar la adecuación de datos, modelos y resultados al uso que pretendemos hacer de ellos, denominado en la literatura anglosajona como fitness for use.

Pretendemos abordar una parte de la validación, mejorando la metodología propuesta por Gómez Delgado y Bosque Sendra (2004), basada en un análisis de incertidumbre, cuyo objetivo es analizar el alcance del efecto que los posibles errores en los datos de partida del modelo puedan provocar en los resultados.

Dicha metodología se basa en la introducción de distorsiones, tanto en la componente espacial, como en la componente temática de los datos de partida. En el trabajo mencionado estas distorsiones se realizan de manera totalmente aleatoria. En la actual propuesta se controla la dimensión y alcance de estas distorsiones, intentando plantear una situación más realista, reproduciendo los errores más probables y descartando aquellos que no podrían haber pasado desapercibidos. Los resultados se evalúan mediante el cálculo de la Incertidumbre Media Cuadrática (IMC), índice que nos dará información sobre el alcance e impacto de estas distorsiones en los resultados del modelo (Gómez Delgado y Bosque Sendra, 2004). Así mismo se compararán dichos resultados con los obtenidos aplicando la metodología original.

De manera resumida podremos decir que, utilizando una metodología u otra, los resultados del modelo son bastante robustos. Pensamos que esta metodología mejora la anteriormente planteada y permite realizar validaciones más realistas.

PALABRAS CLAVE

Análisis de incertidumbre, Sistemas de Información Geográfica, evaluación multicriterio, simulación del crecimiento urbano.

ABSTRACT

A model is always a simplified version of some reality that allows us to describe and understand a problem, collecting the essential elements and mechanisms of the real-world systems. However, a necessary aspect is to demonstrate that these instruments generate reliable representations of the simulated systems through the application of a validation process.

In this paper, the model to validate reproduces a future scenario of urban growth based on Multiple Criteria Evaluation techniques (MCE) and Geographic Information Systems (GIS). The main problem in this topic is to determine the adequacy of data, models and results to their use, called in Anglo-Saxon literature fitness for use.

We aim to carry out a part of that validation, improving the methodology proposed by Gómez Delgado and Bosque Sendra (2004), based on an uncertainty analysis, which analyzes the effect of possible errors from the input data over the results of the model.

This methodology is based on the introduction of distortion in both spatial and thematic component of the input data. In that work, these distortions are made completely at random. In our current proposal, we control the size and scope of the distortions trying to raise a more realistic situation; using the most likely errors and discarding those that could not have go unnoticed. The results are evaluated by the Mean Square Uncertainty (IMC, Incertidumbre Media Cuadrática), index which will give us information about the impact of the distortions into the model results (Gómez Delgado and Bosque Sendra, 2004). Our results are also compared with those obtained using the original methodology.

In summary we can say that, using our methodology or that other, the model results are quite robust. We believe that our methodology improves the above one developed and allows us to make more realistic validations.

KEY WORDS

Uncertainty analysis, Geographical Information Systems, multicriteria evaluation, simulation of urban growth.

1 INTRODUCCIÓN

La aplicación de procesos de validación que aporten cierta robustez y confianza a los resultados obtenidos de los análisis y modelos ejecutados en un entorno SIG sigue siendo una asignatura pendiente. Esta práctica es más importante, si cabe, cuando tratamos de modelos de simulación de la evolución de un fenómeno a futuro, en este caso el crecimiento urbano, para el que no existirán datos reales del sistema que se intenta reproducir con los que poder contrastar los resultados.

El análisis de incertidumbre se plantea aquí como parte de un proceso de validación que se basa en la transmisión de los posibles errores existentes en las variables de entrada del modelo a los resultados del mismo. Teniendo en cuenta esto, una distorsión controlada de las variables iniciales puede proporcionar un índice que valore la incertidumbre y evalúe, de alguna manera, la robustez de los resultados de los modelos, tal y como proponen Gómez Delgado y Bosque Sendra (2004).

Partiendo de la propuesta de estos autores, pretendemos desarrollar una metodología alternativa, que forme parte de una validación parcial de un modelo de simulación de crecimiento urbano deseable para el año 2020 en la Comunidad de Madrid, basado en técnicas de EMC y SIG.

1.1 Planteamiento del problema

En la actualidad se están desarrollando nuevos procedimientos capaces de proporcionar al usuario un resumen lógico sobre la incertidumbre en los resultados de un modelo (Devillers *et al*., 2007), pero, en cualquier caso, no existe un acuerdo general en la comunidad científica sobre en qué debe consistir ese proceso de validación.

Desde los años 90 del pasado siglo, se vienen desarrollando distintos procedimientos para evaluar la calidad de los datos utilizados en un modelo ejecutado en un entorno SIG. Ya en 1987, Bédard (1987) destacó la importancia del estudio de la incertidumbre en las bases de datos.

Por un lado, existen ciertos estudios basados en reglas y procedimientos. Ejemplos como los de Agumya y Hunter (1999), que proponen una metodología de seis fases para la gestión de riesgos en el uso de datos espaciales, o los de Gervais (2003), con manual de instrucciones para desarrollar una estrategia de gestión de riesgos derivados de la incertidumbre.

Otras metodologías están basadas en técnicas de visualización. Así, encontramos

ejemplos como los de Leitner y Buttenfield (2000), que representan la certeza de los datos con visualizaciones de saturación, textura y valor; Howard y MacEachren (1996) o Fisher (1994), emplean simbologías y estudios sobre el esfuerzo que ha de realizar un lector de un mapa para deducir la incertidumbre asociada; Pontius y Schneider (2001) emplean el parámetro Relative Operating Characteristic (ROC).

Existen ciertos análisis de incertidumbre relacionados con los posibles errores en los datos de partida. Ejemplos como los realizados por Emmi y Horton (1996) en modelos para riesgo sísmico; las técnicas empleadas por Davis y Keller (1997) de lógica difusa y simulación de Monte Carlo en el análisis de incertidumbre de un modelo de estabilidad de taludes; los cálculos de Kiiveri (1997) para la distorsión geométrica y análisis de incertidumbre de un modelo de localización óptima para el emplazamiento de una nueva construcción, o los estudios de Arbia y Haining (1998) sobre la propagación de errores en operaciones de superposición mediante el análisis ANOVA y el índice Kappa. Burnicki *et al*., (2007) proponen una metodología basada en la generación de un conjunto de mapas que son perturbados con patrones específicos de error. Acto seguido comparan los mapas generados y determinan los efectos de los patrones de error en la exactitud del mapa resultante. Lee *et al*., (1992) también añaden una perturbación aleatoria para analizar los errores de los modelos.

Así pues, se pueden generar unas variables iniciales afectadas por una distorsión controlada para obtener su modelo derivado y, a continuación, comparar el modelo obtenido con el original para realizar el análisis del error transmitido. Por su parte, Gómez Delgado y Bosque Sendra (2004) proponen una metodología similar que pretende controlar el riesgo en cualquier toma de decisiones a partir de los resultados de un modelo basado en técnicas de EMC y ejecutado en un entorno SIG. Esta metodología parte del uso combinado de un análisis de sensibilidad y un análisis de incertidumbre. Nuestro trabajo desarrolla este análisis de incertidumbre y analiza el efecto que los errores iniciales provocan en los resultados.

Finalmente, hemos de señalar que, en muchos de los casos mencionados, los procedimientos resultan muy complejos y difíciles de implementar, por lo que resulta necesario desarrollar propuestas que permitan ser puestos en práctica con las herramientas disponibles en cualquier SIG convencional y crear módulos de software o actualizar los existentes, para automatizar estos procesos.

2 METOLODOGÍA

2.1 Descripción del modelo a validar

Nuestro trabajo se desarrolla sobre uno de los tres escenarios de crecimiento urbano simulados en el proyecto de investigación SIMURBAN[1], a partir de técnicas EMC y SIG. La metodología aquí planteada se ensaya sobre el escenario de innovación y sostenibilidad, concretamente sobre el modelo de uso residencial desarrollado en la Comunidad Autónoma de Madrid para el año 2020, partiendo de la situación existente en el año 2000, aunque en el proyecto también se llevó a cabo la simulación para los usos del suelo industrial y comercial. Este escenario se materializa a partir de una serie de factores sociales, económicos y ambientales que persiguen conformar una ocupación sostenible del territorio. Esta serie de factores fueron ponderados a partir del método de comparación por pares de Saaty y a partir de una sumatoria lineal ponderada se obtuvo el mapa de aptitud de uso urbano. Es posible encontrar información más detallada de la estructura y resultados del modelo en Plata *et al*., 2011.

2.2 Procedimiento de análisis de incertidumbre

Las fases de aplicación serían:

- En primer lugar se inserta un error aleatorio, en este caso controlado, tanto en la componente espacial como temática de las variables iniciales.
- Se ejecuta de nuevo el modelo, empleando las variables a las que se les ha introducido el error.
- Se repite el proceso del punto uno y dos un número de veces estadísticamente significativo, obteniendo así una serie de resultados sensiblemente diferentes.
- Finalmente, se mide el grado de variación alcanzado mediante un índice que compare los resultados de los modelos modificados con respecto a los resultados del modelo original.

La información resultante sirve para obtener una valoración de la incertidumbre y el grado de confianza de los resultados del modelo.

En nuestro trabajo, se aplicarán diferentes tipos de distorsión en función del tipo de variable que se represente y de las dimensiones de sus

[1] SIMURBAN: Análisis y simulación prospectiva mediante Tecnologías de la Información Geográfica (TIG) del crecimiento urbano actual. Evaluación de su sostenibilidad", (SEJ2007-66608-C04- 00/ GEOG). Es posible encontrar más información en www.geogra.uah.es/simurban/

entidades. Pretendemos introducir errores más próximos a lo que podría ocurrir en la realidad y, además, introducirlos en las variables originales (antes de su modelación y conversión a factor) para reflejar las posibles afecciones del proceso de obtención del factor.

2.2.1 Distorsión geométrica de las variables de partida

La distorsión geométrica se realiza mediante un proceso inverso al de georreferenciación. De esta forma, en lugar de desplazar los elementos a una posición correcta, estos tomarán nuevas localizaciones dentro de un margen de incertidumbre espacial.

En la incertidumbre de un mapa digital, cada conjunto de coordenadas y, dentro de éste, cada coordenada, tiene un grado de aproximación o dispersión que es función de un sinnúmero de causas accidentales o sistemáticas. En relación a esta dispersión, el margen de incertidumbre espacial hace referencia a la magnitud y ámbito de error en el que los elementos espaciales puedan localizarse.

Con respecto a la magnitud del error, supongamos un nodo que represente la localización de un hospital o de un almacén de residuos que tiene una incertidumbre en su posición. Esta magnitud de error, en formato distancia, está aplicada dentro de un radio centrado en la posición real de la entidad. Para aplicar esta magnitud de error, se propone emplear una distribución aleatoria de vectores que distorsionen geométricamente las entidades. En nuestro modelo, las magnitudes de los vectores empleados miden una media de 350m con variaciones entre 0 y 700m. Estos valores aplicarán un desplazamiento medio de 6 píxeles y un máximo de 14, que corresponden a la resolución de 50m empleada en todos los factores de nuestro modelo.

Con respecto al ámbito de distorsión, hacemos referencia a la disposición espacial de cada vector, es decir, la frecuencia y disposición espacial en la que las magnitudes aleatorias van a estar aplicando su efecto. En nuestro modelo, hemos repartido los vectores por todo el territorio de una forma aleatoria y uniforme, evitando aglomeraciones, vacíos y superposiciones. El ámbito, la frecuencia y dimensiones de los vectores se ha diseñado en función del tiempo de procesado y de la potencia del hardware y software empleado. Se han realizado diversas pruebas para nuestro modelo y se ha optado finalmente por una matriz virtual de celdas de dimensiones de 2 por 2 kilómetros, separadas entre ellas 500 metros, en cada una de las cuales se colocará un vector (figura 1). Cubrir toda la imagen de estudio con esta distribución ha supuesto alrededor de tres mil vectores.

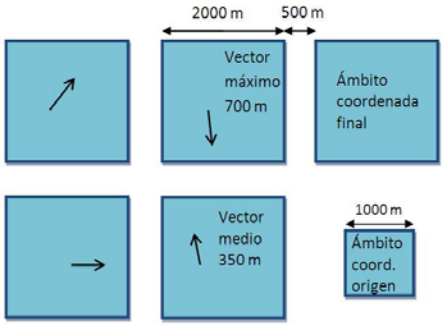

Figura 1. Geometría de disposición de los vectores identidad en la distorsión espacial. Cada vector tiene el punto origen y final dentro de un ámbito de posición.

El carácter aleatorio lo proporciona el empleo de números aleatorios obtenidos por ordenador. Así se obtienen valores diferentes para cada par de coordenadas x e y en cada versión del factor distorsionado. El error E a introducir en cada coordenada se puede obtener a partir de la expresión:

$$E_{i,m,n,j} = M_i \cdot A_{i,m,n,j} \cdot (-1)^{entero(100 \cdot A_{i,m,n,j})} \quad (1)$$

$\forall\, i = \{x, y\}\ ,\ j = \{o, f\}$. Donde *M* es la semi-magnitud del error máximo para cada eje de coordenadas x o y; m y n corresponden a la fila y columna de la matriz de celdas virtuales; i indica si es para la coordenada x o y; j indica si es el punto origen (o) o punto final (f) del vector y A es un número entero aleatorio entre 0 y 1.

Entonces, la coordenada origen y final para cada vector aleatorio será:

$$\begin{aligned} x_{m,n,o} &= x_{\emptyset} + n \cdot d_x + E_{x,m,n,o} \\ y_{m,n,o} &= y_{\emptyset} + m \cdot d_y + E_{y,m,n,o} \\ x_{m,n,f} &= x_{\emptyset} + n \cdot d_x + E_{x,m,n,o} + E_{x,m,n,f} \\ y_{m,n,f} &= y_{\emptyset} + m \cdot d_y + E_{y,m,n,o} + E_{y,m,n,f} \end{aligned} \quad (2)$$

Donde sólo queda definir xØ e yØ como las coordenadas de la esquina del mapa y dx y dy como los incrementos de distancia entre cada vector, que se pueden obtener dividiendo la longitud del mapa entre el número de vectores que se quiera introducir por cada eje x o y. El último término está en forma de potencia y cambia aleatoriamente el signo.

Con este concepto geométrico, aplicaremos una distorsión espacial empleando una transformación bidimensional. La metodología

propuesta por Gómez Delgado y Bosque Sendra (2004) se basa en transformaciones polinómicas, disponibles en la mayoría de SIG comerciales, que distorsionan los elementos espaciales mediante funciones de interpolación. En este trabajo consideramos que esta distorsión resulta insuficiente para el propósito planteado, pues depende del número de coeficientes empleados en los polinomios, que son resultado del promedio medio cuadrático de todos los vectores identidad. Así planteamos acudir a otro tipo de transformación, por distorsión, que se fundamente en la existencia de un patrón de comportamiento. Modelar este tipo de transformación puede llevarse a cabo por técnicas de superficies de mínima curvatura, por mínimos cuadrados, por rubbersheet o por regresión múltiple, entre otras (González Matesanz *et al*., 2004). En el software ArcGis 9.2 se puede encontrar la herramienta de ajuste por distorsión rubbersheet, que proporciona deformaciones más locales que los ajustes polinómicos y la posibilidad de moldear mejor el concepto geométrico establecido. El proceso rubbersheet se fundamenta en encontrar una transformación entre coordenadas originales y distorsionadas, mediante una triangulación de Delaunay, determinando un homomorfismo entre las dos partes (Saafeld, 1985; White y Griffin, 1985). En la figura 2 podemos observar los efectos de la transformación rubbersheet sobre las entidades del mapa de tipo de suelos. Para poder apreciar las distorsiones que esta transformación proporciona se han superpuesto los contornos de las entidades del mapa original. Es posible encontrar información más detallada sobre la estructura y resultados de esta transformación en Calabia Aibar (2011).

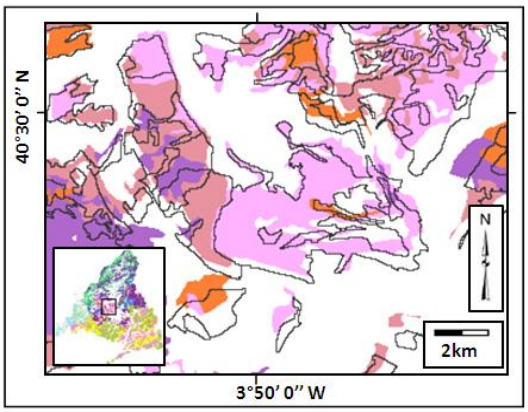

Figura 2. Efecto de la distorsión geométrica en una zona del mapa de tipo de suelos. Los contornos originales están superpuestos para observar que su distorsión se realiza en todas direcciones y con diferentes magnitudes.

2.2.2 Alteración del Modelo Digital de Elevaciones (MDE) para la distorsión de los factores pendiente y orientación.

En el modelo analizado se emplea un MDE para generar los factores pendiente y orientación del terreno. Gómez Delgado y Bosque Sendra (2004) proponen alterar este tipo de factores tratándolos como una variable temática e introduciendo la perturbación a nivel de píxel. Sin embargo, parece razonable pensar que debería alterarse el MDE original del que se derivan.

Así pues, siendo el MDE una matriz regular a la que se puede introducir una perturbación, la idea es sumar al MDE un modelo digital de superficie de error aleatorio que lo altere. Lee *et al*., (1992) simulan esta magnitud de error mediante un generador de números aleatorios. Los números aleatorios generados se vuelven a escalar para estar dentro de una magnitud especificada y luego son cambiados y controlados para tener varios grados de clústeres espaciales. Esta autocorrelación espacial, Lee *et al*., (1992) la realizan mediante el índice de Moran (Goodchild, 1986). Nosotros proponemos emplear una simple red de triángulos irregulares (TIN del inglés Triangular Irregular Network) de magnitud y distribución controlada que, tras revisar ciertos estudios relacionados con el error en los MDE (Kubik y Wu, 1995; Mukai *et al*., 1989; Priebbenow y Clerici, 1988; Sasowsky *et al*., 1992 o Tahiri y Donnai, 1995) y teniendo en cuenta la resolución empleada en el modelo (50x50m), provoca un máximo de variación en altura de ±5 metros y un máximo de variación de pendiente del ±10%. Para obtener dicha red, se creará un punto de cota, aleatorio en planimetría y altimetría, en una matriz virtual similar a la de la figura 1, de celdas cuadradas de 400 metros de lado y separadas entre ellas 100 metros. Con estas dimensiones se ha conseguido un centenar de miles de puntos para toda la imagen analizada. Las coordenadas se pueden obtener fácilmente con lenguajes de programación o con una simple hoja de cálculo, empleando ecuaciones similares a (1) y (2). Una vez creada la TIN de error, sólo queda sumar el MDE.

2.2.3 Distorsión temática de las variables de partida.

La distorsión temática pretende simular la incertidumbre de la clasificación de las entidades de un mapa.

El proceso de distorsión empleado por Gómez Delgado y Bosque Sendra (2004) se realiza sobre los píxeles que componen la imagen ráster que

representa a cada uno de los factores incluidos en el modelo. Así, comienza estableciendo el volumen de píxeles a distorsionar de un 4 a 8% del área de estudio. Para variables de tipo cualitativo, se lleva a cabo un cambio de categoría aleatorio, asignando nuevos identificadores a cada uno de los píxeles. Para las variables de tipo cuantitativo se modifica un porcentaje en función del valor máximo registrado de cada factor original con valores extraídos de una tabla de números aleatorios. Estos píxeles con "ruido" crean un efecto de "sal y pimienta", ya que son seleccionados de manera aleatoria sobre todo el área.

La distorsión que aquí proponemos pretende cambiar aleatoriamente los atributos cualitativos asignados para cada entidad y no para píxeles singulares. Éstos cambiarán teniendo en cuenta el atributo original de dicha entidad y el error temático asociado a éste. Poniendo un ejemplo, es menos probable que una superficie de agua se haya confundido con cualquier otro tipo de cubierta, que un suelo descubierto se haya confundido con un cultivo de secano. De este modo, en lugar de realizar variaciones a nivel píxel, proponemos realizar una perturbación dirigida, que varíe los atributos cualitativos de cada identidad en formato vectorial original (punto, línea o polígono). Para cuantificar el error temático asociado al atributo cualitativo se utilizarán las matrices de confusión que son empleadas habitualmente para verificar la fiabilidad de la clasificación de las entidades (Chuvieco, 2008). Así, las celdas de estas matrices indicarán el valor probabilístico de la exactitud de los atributos asignados y de sus posibles asignaciones erróneas. Las filas indicarán los atributos sin distorsionar y las columnas los resultantes tras la distorsión. La diagonal principal indicará la cantidad de entidades a las que no se les aplica distorsión y el resto de valores indicará la cantidad de entidades a las que sí que se les aplica (Calabia, 2011).

Otra innovación a la metodología de Gómez Delgado y Bosque Sendra (2004) es que, además, pretendemos incorporar restricciones a determinados cambios en función de las características de cada entidad. De esta manera, se evita cambiar el atributo de entidades que superen una determinada área, longitud u otra característica, pues es poco probable que un polígono urbano, tal y como es el de Madrid, cambie erróneamente de clasificación, ya que dicho error hubiese sido detectado y corregido. Para polígonos, por ejemplo, sólo se modificarán aquellos que no superen 1.000 hectáreas, permitiendo, por otro lado, la distorsión de una porción razonable de toda el área de estudio. De la misma forma, se evitará realizar cambios cuantitativos poco probables, tal y como se realiza en la distorsión del volumen de población, donde cada núcleo urbano se modifica de manera controlada y no superando unas cantidades que hubiesen sido fácilmente detectadas, pues es poco probable que si el núcleo urbano de Madrid apareciese con un volumen de población de cinco mil habitantes, este error hubiese pasado desapercibido.

Para implementar esta distorsión se puede emplear una hoja de cálculo o algún lenguaje de programación con algoritmos como los desarrollados por Calabia Aibar (2011). El empleo de números aleatorios es de gran utilidad a la hora de introducir probabilidades de cambio temático, pues se realiza mediante un algoritmo condicional que asigna los posibles atributos en función del valor aleatorio obtenido. Poniendo un ejemplo, si la matriz de confusión indica que la probabilidad de que un atributo esté bien asignado es de un 94%, un valor aleatorio (normalizado de 0 a 1) que esté entre 0 y 0,94 no permitirá asignar otro atributo a la entidad a alterar, pero si está entre 0,94 y 1, el atributo se alterará y cambiará a la otra clase que indique dicha matriz de confusión. Es posible encontrar información más detallada sobre los algoritmos y las matrices de confusión empleadas para esta distorsión en Calabia Aibar, 2011.

En las siguientes secciones se describe la distorsión temática realizada a cada uno de los factores incluidos en el modelo original, habiendo sido descartada la distorsión del factor vulnerabilidad a la contaminación de los mantos acuíferos, porque todos polígonos tienen una gran extensión y, por tanto, cualquier error hubiera sido detectado con facilidad. Una vez distorsionada cada variable original, se ha elaborado el factor pertinente, otorgando los pesos establecidos en el modelo original correspondientes a cada categoría y procediendo posteriormente a su normalización.

a) Distorsión temática del factor tipo de suelos: Con este factor se intenta que las nuevas zonas urbanas no sean asignadas, hasta donde sea posible, sobre los suelos de mayor productividad agrícola. En este caso, además de la matriz de confusión del mapa tipo de suelos, en los algoritmos se incluyeron sentencias condicionales que permitieron la distorsión en función de si el área de cada polígono superase las 1.000 hectáreas. La figura 3 muestra un ejemplo gráfico del resultado de la distorsión temática, para este factor, en una región de la zona de estudio.

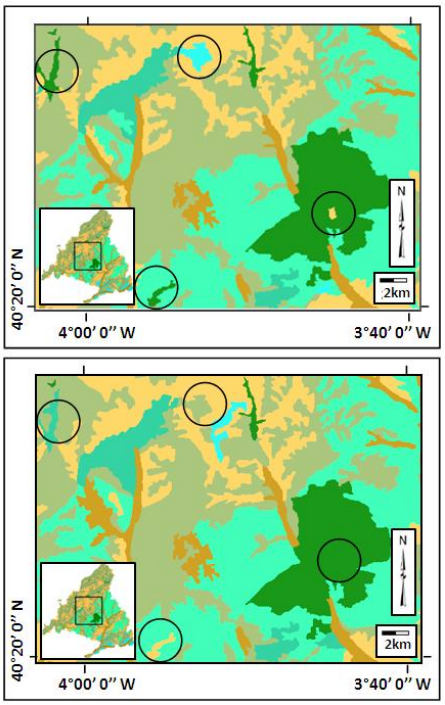

Figura 3. Dos versiones del mapa de aptitud según tipo de suelos tras la distorsión temática de una región de la zona de estudio.

b) Distorsión temática del factor usos del suelo: Este factor considera la protección de las zonas agrícolas, los pastizales y otros usos del suelo con valor ambiental y económico. La matriz de confusión empleada se ha obtenido a partir del estudio sobre errores detectados en la base de datos de usos del suelo CORINE *Land Cover* al realizar un análisis de cambios de usos del suelo en la Comunidad de Madrid entre el año 2000 y 2006 (Barreira, 2011). Como en el caso anterior, también se aplicó la distorsión a los polígonos menores de 1.000 hectáreas. Con esta restricción aplicada a la distorsión dirigida por matriz de confusión, la superficie distorsionada total ascendió a un 15 % del área de la Comunidad de Madrid.

c) Distorsión temática del factor distancia a carreteras: Este factor se incluye en el modelo con la finalidad de que las nuevas zonas urbanas tengan buena accesibilidad. La matriz de confusión aplica la distorsión temática a la variable de partida (mapa de carreteras) intercambiando los identificadores de las vías más probables de ser confundidas. Así, resultaría poco probable haber confundido una autovía asignándole el identificador de una travesía, pero sí podríamos encontrarnos un camino como calle o travesía, o una calle identificada como travesía.

d) Distorsión temática del factor geotecnia: Este factor representa el riesgo de fallas activas, riesgo sísmico o sitios de baja estabilidad geomorfológica. Aquí también se aplicó la distorsión sólo a los polígonos menores de 1.000 hectáreas. Finalmente sólo un 7 % del área de la Comunidad de Madrid fue sometida a la distorsión temática.

e) Distorsión temática del factor distancia a zonas urbanas: Este factor pretende que las nuevas zonas urbanas estén cerca de otros núcleos de población de tamaño intermedio. En este caso, la distorsión temática se realiza mediante varios algoritmos que introducen cambios en el tamaño de la población. Estos algoritmos cambian el tamaño de las ciudades ponderadamente. Para ello, la siguiente fórmula modifica los valores de población:

$$Pob_{mod} = Pob + S \cdot A \cdot (-1)^{Entero} \quad (3)$$

Donde *Pob_mod* es el tamaño de la población modificado; *A* es un número aleatorio entre 0 y 1 y *S* toma diferentes valores en función del tamaño de la población. Los valores empleados para esta variable *S* fueron los siguientes:

i. S = 10 000, si la población es menor a 10 000 hab.
ii. S = 20 000, si está entre 10 000 y 50 000 hab.
iii. S = 40 000, para el resto.

f) Distorsión temática del factor distancia a hospitales: Este factor pretende aportar cierta sostenibilidad desde el punto de vista social, facilitando el acceso de la población a los servicios básicos, concretamente minimizando la distancia a centros sanitarios. Para su distorsión temática se eliminó aleatoriamente el 15% de los hospitales existentes.

2.3 Ejecución del modelo distorsionado y evaluación de los resultados a partir del cálculo de la Incertidumbre Media Cuadrática (IMC)

Una vez distorsionados temática y geométricamente cuatro veces todos los factores originales, se ejecutó el modelo treinta veces, escogiendo aleatoriamente los factores de entre los cinco disponibles, es decir, los cuatro distorsionados más el original. Para evaluar el impacto de los posibles errores que pudiese tener la base de datos utilizada en el modelo sobre los resultados del mismo, se procedió a calcular para cada píxel la IMC, índice desarrollado por Gómez Delgado y Bosque Sendra (2004):

$$IMC = \sqrt{\frac{\sum_0^n (X_i - X)^2}{n-1}} \quad (4)$$

Donde Xi es el valor de adecuación alcanzado por cada píxel en cada una de las 30 ejecuciones del modelo; X es el valor de adecuación

alcanzado por cada píxel en la ejecución del modelo original y n el número de ejecuciones realizadas (en nuestro caso 30).

3. RESULTADOS Y COMPARACIÓN CON LOS RESULTADOS DE LA METODOLOGÍA ORIGINAL

Los valores de IMC correspondiente a los polígonos de asignación óptima se mantienen por debajo del 33%, tienen una media de 6% y una desviación estándar de 4%, pudiendo considerarse, por tanto, que los resultados del modelo original son bastante robustos.

La metodología empleada se ha contrastado con la propuesta por Gómez Delgado y Bosque Sendra (2004). Para ello, se ha llevado a cabo el proceso propuesto por estos autores de distorsión espacial y temática con el modelo objeto de estudio (Benavides, 2011). Posteriormente, se ejecutó el modelo también 30 veces y se calculó la IMC. En la figura 4 se comparan gráficamente los resultados obtenidos siguiendo las dos metodologías para una misma zona del área de estudio. La metodología propuesta arroja resultados más uniformes que la contrastada, que muestra valores de IMC altos en los contornos. Asimismo, los histogramas para ambas metodologías demostraron una distribución más normal para la nueva metodología empleada.

Finalmente, con los polígonos residenciales óptimos obtenidos con el modelo original, se calculó el valor de IMC medio para cada polígono y se creó el mapa de la figura 5, que muestra los polígonos residenciales óptimos que tienen un índice de IMC medio inferior al 6% y, por tanto, son las soluciones óptimas, fiables y robustas.

4 CONCLUSIONES

En el presente trabajo se ha desarrollado una metodología alternativa para el análisis de la incertidumbre que en los resultados de un modelo basado en técnicas EMC y SIG pudieran introducir los posibles errores en los datos de partida. Ante los resultados obtenidos, queda demostrado que el empleo de este tipo de análisis es una herramienta útil para la validación de estos modelos y para confirmar la robustez de los resultados derivados de los mismos.

La metodología propuesta dirige y controla las perturbaciones a la hora de distorsionar los factores empleados en los modelos basados en técnicas de EMC y SIG.

En cuanto a la distorsión espacial, queda demostrado que el empleo de las transformaciones polinómicas no controla la magnitud ni el ámbito de las distorsiones aplicadas y que una transformación adecuada sería la rubbersheet.

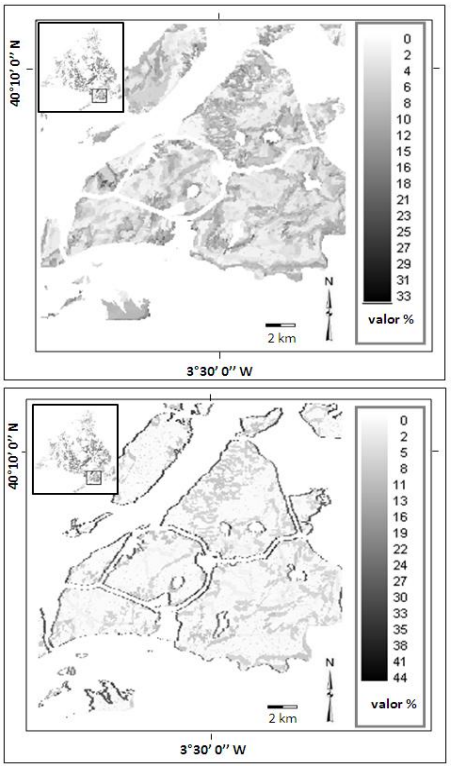

Figura 4. Mapa superior: índice de IMC obtenido en este trabajo para una zona en concreto. Mapa inferior: índice de IMC obtenido para la misma zona pero por distorsión con polinomios cúbicos y ruido temático.

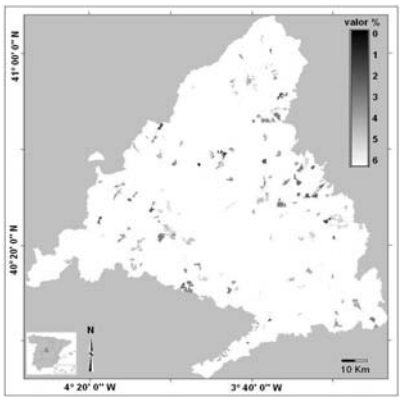

Figura 5. Polígonos residenciales óptimos obtenidos en el modelo original, que han registrado una IMC media inferior al 6% en el proceso de validación mediante análisis de incertidumbre.

En cuanto a la metodología empleada para la distorsión de los modelos digitales derivados, la creación de la superficie de error para

distorsionar los MDE es un proceso más realista que el propuesto por Gómez Delgado y Bosque Sendra (2004), puesto que nos permite reproducir mejor errores reales de las bases de datos iniciales. Además, esta superficie se puede crear fácilmente a partir de una red de triángulos irregulares aleatorios, cosa que simplifica mucho las propuestas de otros autores. Con respecto a la metodología empleada en la distorsión temática, los análisis realizados muestran que el empleo de la matriz de confusión para la introducción de error puede ser también un método más apropiado para la transmisión de la incertidumbre temática.

Finalmente, se puede concluir que aplicar el índice de IMC en los mapas de aptitud para la asignación final del uso del suelo modelado, favorece la realización de una toma de decisiones más informada, fiable y robusta.

AGRADECIMIENTOS

Este artículo se ha elaborado dentro de las actividades del proyecto de investigación de I+D-i: SIMURBAN: Análisis y simulación prospectiva mediante Tecnologías de la Información Geográfica (TIG) del crecimiento urbano actual. Evaluación de su sostenibilidad, financiado por el Ministerio de Educación y Ciencia (referencia SEJ2007-66608-C04- 00/ GEOG).

BIBLIOGRAFÍA

Agumya, A. y Hunter, G.J. 1999. A risk-based approach to assessing the 'fitness for use' of spatial data, *URISA Journal*, 11, 33-44.

Arbia, G. y Haining, R. 1998. Error propagation modelling in raster GIS: overlay operations, *International Journal of Geographical Information Science*, 12, 2, 145-167.

Barreira, P. 2011. *Metodología aplicada a la detección de errores temáticos en el CORINE Land Cover a través del estudio de cambios: Comunidad de Madrid (2000-2006)*, Proyecto fin de Carrera, Universidad de Alcalá.

Bédard, Y. 1987. Uncertainties in land information systems databases, *Proceedings of Eighth International Symposium on Computer-Assisted Cartography.* Baltimore, 175–184.

Benavides Gutiérrez, G. 2011. *Análisis de incertidumbre de un modelo basado en técnicas SIG y EMC,* Proyecto fin de Máster, Universidad de Alcalá.

Burnicki, A.C., Brown, D.G. y Goovaerts, P. 2007. Simulating error propagation in land-cover change analysis: the implications of temporal dependence. *Computers, Environment and Urban Systems*, 31 (3), 282-302.

Calabia, A. 2011. *Metodología para el análisis de incertidumbre de un modelo basado en técnicas de Evaluación MultiCriterio y Sistemas de Información Geográfica*, Proyecto fin de Carrera, Universidad de Alcalá. http://comunidad.calabia.com/documentos/2011.pdf

Chuvieco, E., 2008. *Teledetección Ambiental: La observación de la Tierra desde el espacio,* p. 501.

Davis, T.J. y Keller, C.P. 1997. Modelling uncertainty in natural resource analysis using fuzzy sets and Monte Carlo simulation: slope stability prediction, *International Journal of Geographical Information Science*, 11, 5, 409-434.

Devillers, R., Bédard, Y., Jeansoulin, R. y Moulin, B. 2007. Towards spatial data quality information analysis tools for experts assessing the fitness for use of spatial data, *International Journal of Geographical Information Science*, 21 (3-4), 261-282.

Emmi, P.C. y Horton, C.A. 1996. Seismic risk assessment, accuracy requirements, and GIS-based sensitivity analysis. En Goodchild, M.F, Steyaert, L.T. Y Parks, B.O. (Eds.): *GIS and Environmental modeling: progress and research issues*, Fort Collins, CO: GIS World Books, pp. 191-195

Fisher, P.F., 1994. Animation and sound for the visualization of uncertain spatial information. En *Visualization in Geographic Information Systems*, H.M. Hearnshaw y D.J. Unwin (Eds), Chichester, UK: Wiley, pp. 181–185.

Gervais, M. 2003. *Pertinence d'un manuel d'instructions au sein d'une stratégie de gestion du risque juridique découlant de la fourniture de données géographiques numériques*, PhD thesis, Sciences Géomatiques, Québec, Université Laval.

Gómez Delgado, M. y Bosque Sendra, J. 2004. Aplicación de análisis de incertidumbre como método de validación y control del riesgo en la toma de decisiones, *GeoFocus*, 4, 179-208.

González Matesanz, F.J., Quiros, R., Cano, M., Sánchez Sobrino, J. y Dalda, A. 2004. Cambio de datum ED50 – ETRS89. Métodos y resultados. *Congreso Nacional de Topografía y Cartografía TOPCART*.

Goodchild, M.F. 1986. Spatial Autocorrelation, Catmog 47, *Geo Books*, Norwich.

Howard, D. y Maceachren, A.M. 1996. Interface design for geographic visualization: Tools for representing reliability, *Cartography and Geographic Information Systems*, 23, 59–77.

Kiiveri, H.T. 1997. Assessing, representing and transmitting positional uncertaintiy in maps, *International Journal of Geographical Information Science*, 11 (1), 33-52.

Kubik, K. y Wu X. 1995. Mapping from SPOT images using digital photogrammetric workstation, *17th International Cartographic Conference Proceddings*, Barcelona, 1, 266-274.

Lee, J., Snyder, P. K. y Fisher, Y P. F. 1992. Modeling the effect of data errors on feature extraction from digital elevation models, *Photogrammetric*

Engineering and Remote Sensing, 58(10): 1461-1467

Leitner, M. y Buttenfield, B.P. 2000. Guidelines for the Display of Attribute Certainty, *Cartography and Geographic Information Science*, 27, 3–14.

Mukai, Y., Sugimura, T., Iijima, T., Ayabe, K., Tachi, K. y Nakazawa, T. 1989. Automatic computation of elevation data of side-lap area using system corrected Landsat TM data, *Technical Report. Remote Sensing Technology Center of Japan*. Tokyo.

Munroe, D., Southworth, J. y Tucker, C.M. 2002. The dynamics of land-cover change in western Honduras: Exploring spatial and temporal complexity, *Agricultural Economics*.

Plata Rocha, W., Gómez Delgado, M. y Bosque Sendra, J. 2011. Simulating urban growth scenarios using GIS and Multicriteria Evaluation techniques: a case study of the Madrid Region, Spain, *Environment and Planning B: Planning and Design*, Vol.38, 1012-1031.

Pontius, R. G., y Schneider, L. C. 2001. Land cover change model validation by an ROC method for the Ipswich watershed. Massachusetts, USA, *Agriculture, Ecosystems and Environment*, 85, 239–248.

Priebbenow, R. y Clerici, E. 1988. Cartographic applications of SPOT imagery, *International Archives of Photogrammetry*, 37 (Part B4), 289-297.

Rykiel, E.J. JR. 1996. Testing ecological models: the meaning of validation, *Ecological Modelling*, 90, 229-244.

Saafeld, A. 1985. A fast Rubbersheeting transformation using simplicial coordinates, *The American Cartographer*, 12,169-173.

Sasowsky, K.C., Petersen, G.W. y Evans, B.M. 1992. Accuracy of SPOT digital elevation model and derivatives: Utility for Alasks's North Slope, *Photogrammetric Eng. Remote Sens*., 58, 815-824.

Tahiri, D. y Donnay, J-P. 1995. Modélisation de la précision des modèles numériques de terrain dérivés des cartes topographiques. *Proceedings of the 17th International Cartographic Conference of the ICA*, Barcelona, 1, 469-473

White, M.S. y Griffin, P. 1985. Piecewise Linear Rubber-Sheer Map Transformation, *The American Cartographer*, 12, 123-131